Elsevier Editorial System(tm) for Computer Communications
Manuscript Draft

Manuscript Number:

Title: Information delivery in delay tolerant networks

Article Type: Review / Tutorial Paper

Keywords: DTN, routing and forwarding, stochastic and deterministic routing, network coding, hybrid routing/encoding

Corresponding Author: Dr. Dmitry Moltchanov, Ph.D.

Corresponding Author's Institution: Tempere University of Technology

First Author: Mozhdeh Gholibegi, MSc.

Order of Authors: Mozhdeh Gholibegi, MSc.; Morteza Karimzadeh, MSc.; Dmitry Moltchanov, Ph.D.; Yevgeni Koucheryavy, Ph.D.

Abstract: Modern Internet protocols demonstrate inefficient performance in those networks, where the connectivity between end nodes has intermittent property due to dynamic topology or resource constraints. Network environments, where the nodes are characterized by opportunistic connectivity are referred to as Delay Tolerant Networks (DTNs). Because of numerous applications such as low-density mobile ad hoc networks, command/response military networks and wireless sensor networks, DTNs have been one of the growing topics of interest characterized by a significant amount of research efforts invested in this area over the past decade. Routing is one of the major components affecting the overall performance of DTN networks in terms of resource consumption, data delivery and latency. Over the past few years a number of routing protocols have been proposed. The focus of this paper is on description, classification and comparison of these protocols. We discuss the state-of-the-art routing schemes and methods in opportunistic networks and classify them to two main categories: deterministic and stochastic routing. Finally, we introduce the recently proposed network coding concept, and discuss recent researches regarding coding-based routing protocols in intermittent networks. We believe that our work would be helpful for those looking for recent developments in the area of routing and forwarding in DTN systems.

Suggested Reviewers:

Opposed Reviewers:

March 07, 2011

Dear Editor of the Computer Communications journal,

Enclosed, there is a paper entitled "**Information delivery in delay tolerant networks**" Please, accept it as a candidate for publication in the **Computer Communications journal**. I give you my assurance that it is not under consideration elsewhere.

If required, I can be contacted using email: moltchan@cs.tut.fi

Thank you for your consideration,

Dr. Dmitri Moltchanov
Department of Communications Engineering,
Tampere University of Technology,
E-mail: moltchan@cs.tut.fi

Information delivery in delay tolerant networks

M. Karimzadeh[1], M. Gholibegi, D. Moltchanov, Y. Koucheryavy
Department of Communications Engineering
Tampere University of Technology
E-mail: morteza.karimzadeh@tut.fi

Abstract. Modern Internet protocols demonstrate inefficient performance in those networks, where the connectivity between end nodes has intermittent property due to dynamic topology or resource constraints. Network environments, where the nodes are characterized by opportunistic connectivity are referred to as Delay Tolerant Networks (DTNs). Because of numerous applications such as low-density mobile ad hoc networks, command/response military networks and wireless sensor networks, DTNs have been one of the growing topics of interest characterized by a significant amount of research efforts invested in this area over the past decade. Routing is one of the major components affecting the overall performance of DTN networks in terms of resource consumption, data delivery and latency. Over the past few years a number of routing protocols have been proposed. The focus of this paper is on description, classification and comparison of these protocols. We discuss the state-of-the-art routing schemes and methods in opportunistic networks and classify them to two main categories: deterministic and stochastic routing. Finally, we introduce the recently proposed network coding concept, and discuss recent researches regarding coding-based routing protocols in intermittent networks. We believe that our work would be helpful for those looking for recent developments in the area of routing and forwarding in DTN systems.

Keywords: DTN, opportunistic networking, routing protocols, network coding

1. Introduction

Delay tolerant networks (DTNs) represent a class of networks, where no assumption regarding the existence of a well-defined path between two nodes wishing to communicate is made. Particularly, source and destination systems might never be connected to the network at the same time and connections among wireless nodes are temporal. Such networks may have

[1]Corresponding authors, e-mail: morteza.karimzadeh@tut.fi

sparse node densities, with short communication capabilities of each node. One-hop connections are often disrupted due to node mobility, energy conservation or interference. However, in these networks, a link can still be established when two nodes come into the coverage range of each other. DTN concept stipulates that such temporal links can be used to exchange information possible on behalf of other nodes hoping that it will eventually reach the destination. Although, this communication paradigm usually involves a lot of overhead in terms of additional delay as packets are often buffered in the network, this seems to be the only viable solution for such specific environments.

The existing TCP/IP based Internet, operates assuming end-to-end communication using a concatenation of various data-link layer technologies. The set of rules specifying the mapping of IP packets into network-specific data-link layer frames at each router provides the required level of interoperability. Still IP protocol makes a number of key assumptions regarding lower layer technologies making seamless IP layer communications smooth. These are: (i) there is an end-to-end path between two communicating end systems, (ii) the round-trip time between communicating end systems is not absurdly high and (iii) the end-to-end packet loss probability is rather small. Unfortunately, in DTN networks one or more of the above mentioned assumptions are violated due to mobility, power conservation schedule or excessive bit error rate. As a result, classic protocols of TCP/IP protocol stack are not appropriate for such environments [21]. A key reason why end-to-end communication is difficult in DTN topology is that IP packet delivery works only when the end-to-end path is available. In practice, according to classic IP routing mechanisms an IP packet is dropped at the intermediate system where no link to the next hop currently exists. Such design restricts the end-to-end communication to those scenarios, where intermediate nodes have to buffer received packets to deliver them whenever they have an opportunity to contact their destinations [9].

This paper discusses the state-of-the-art research in information delivery in DTNs paying special attention to recent developments in the area, e.g. usage of network coding. An interested reader can find older surveys in [16,21,22]. The rest of the paper is organized as follows. In Section 2 we provide a brief overview of DTN architectures, consider their characteristics, and discuss some applications. It is followed by a description of characteristics and challenges in DTNs in Section 3. Then, in Section 4, we describe various routing based information delivery mechanisms in networks with opportunistic connectivity and cover several representative solutions. We also discuss special solutions such as those based on combining routing and forwarding into a single bundle e.g. network coding. Finally, we present some novel open concepts which could potentially become focused research areas further in this field. Particularly, we discuss hybrid approaches and the possibility to use routing and network coding based methods jointly to improve performance of information delivery.

2. Architecture and examples

DTNs represent a unique wireless network architecture enabling mobile nodes to have communication with each other in environments, where there is no continual route between the end nodes. DTNs are alternative structures to traditional networks facilitating connectivity of systems and network regions with sporadic or unstable communication links. In networks with such circumstances, mobile relay nodes are used to carry and forward of messages and make communication possible among other nodes. Depending on DTNs types communication opportunities could be either scheduled over time or completely random.

Opportunistic networks often arise as a result of host and router mobility. Another reason for sporadic connectivity is disconnection due to power management or interference. Some examples are discussed below.

Inter-planet satellite communication network. The DTN protocol emerged from work first started in 1998 in partnership with NASA's Jet Propulsion Laboratory. The initial goal was to modify the TCP protocol to facilitate communications between satellites. The objective of the Interplanetary Internet project was to define the architecture and protocols for interoperation of the Internet resident on Earth with other remotely located residents on other planets or spacecrafts. While the Earth's Internet is basically a "network of interconnected networks", the Interplanetary Internet may therefore be thought of as a "network of disconnected Internets". Internetworking in such environment requires new techniques to be developed [9]. The speed-of-light delays, sporadic and unidirectional connectivity, as well as high bit error rates make the use of current Internet protocols infeasible across such long distances [1].

Sparse mobile ad hoc networks. In many cases, these networks may have unexpectedly intermittent connections due to node mobility or sparse deployment. Sometimes sporadic connectivity in the network could be periodic or predictable. For example, a bus carrying a computer can act as a store and forward message switch with limited RF communication capability. As it travels, provides a form of message switching service to its nearby clients to communicate with distant parties it will visit in the future [9].

Country-side area network. DTNs can bring digital connectivity to rural areas and other environments with limited or non-existing infrastructures. In these networks transportation systems such as cars, buses, and boats are used to provide relaying of messages by moving around and collecting/delivering messages from/to various nodes. Recently, a number of projects have been launched to explore such communication concept. One example is the Message Ferry project aimed at developing a data delivery system in areas without existing Internet infrastructure [36]. Another example is the DakNet project that should potentially provide low-cost connectivity to the Internet to villages in India [3].

Military battlefield network. In a military setting DTN allows for a rich set of applications including the dissemination of mission-critical information in battlefield scenarios. For these types of applications, the delay tolerant protocol should transmit messages across multiple-hop networks consisting of different sub networks based on network parameters such as delay and loss. Such systems may be expected to operate in hostile environments where mobility, noise, attacks, interference and/or intentional jamming may easily result in disconnection and data traffic may have to wait several seconds or more to be delivered [9,1].

Wireless sensor networks. Wireless sensor networks are often characterized by limited end-node resources including energy, memory and CPU power. Communication within these networks is often aimed at limited usage of these resources. Scarcity of power calls for advanced power saving schedules naturally leading to **i**ntermittent connections between nodes. In this scenario, utilization of opportunistic communications becomes very important. Lack of infrastructure may force sensor network gateways to be intermittently connected to operator's network. Scheduled down time, interference, or environmental hostility may cause the interruption of operable communication links [3].

Exotic media networks. Exotic communication media includes near-Earth satellite communications, very long-distance radio links, acoustic transmission in air or water, free-space optical communications and nano-networks [25]. Depending on a certain scenario these systems might be subject to high latencies with predictable or sporadic interruptions (e.g. due to planets' movements or scheduled arrival of a ship/vehicle), may suffer outage due to environmental conditions (e.g. weather), or may even provide a predictably available store-and-forward network service that is only occasionally available (e.g. satellites).

3. DTN characteristics

To discuss the routing and forwarding problem, we need a model capturing the most important characteristics of a DTN network. This section explores them concentrating on those producing the most impact on routing and forwarding protocols, e.g. path properties, network architectures, and end node resource constraints.

Intermittent connection. One of the most important characteristics of DTNs is that the end-to-end connection between communicating end systems may not exist. Generally, intermittent connections may be broadly categorized as due to a fault or not. Non-faulty disconnections happen in wireless environments and mostly caused by two sources: mobility and short duty cycle of system operation. Intermittent connection as a result of mobility depends on the application area of DTNs. Communication schedules can be created based on predictability or can be fully opportunistic. In the latter case nodes come to the coverage area of each other due to their random movement or due to movement of other objects [12,16]. Fig. 1 demonstrates predictability of communication schedules for mobile nodes in different scenarios.

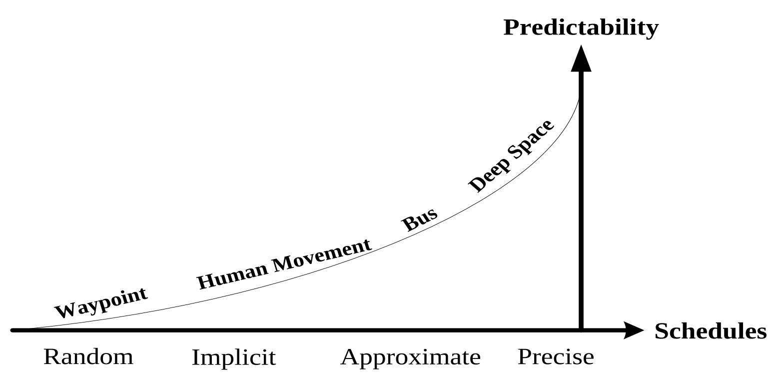

Figure 1. A range of predictability for communication schedule.

Intermittent connections caused by short duty cycles are common among devices with limited resources. These connections are often predictable. Dealing with disconnections requires the routing protocol to “understand” that the lack of connectivity happens as a result of a normal situation rather than force majeure, and should not be considered as a faulty operation [16, 22].

Delivery latency and low data rate. Delivery latency is the amount of time between message injection into the network and its successful reception at the destination. Since many applications can benefit from short delivery times, latency is one of the most important performance metrics of interest. This delay consists of transmission, processing, propagation time over all links as well as queuing delay at each system along the path. In DTNs, transmission rates are often relatively small and latencies can be large. Additionally, data transmission rates can also be largely asymmetric in uplinks and downlinks [21]. In some application scenarios (e.g. deep-space communications), delivery latency may vary from a few minutes to hours or even days and a significant fraction of messages may not be delivered at all. For effective operation over DTNs with high latencies and low link rates, the key is to design the routing protocols and forwarding algorithms matching the actual mobility patterns [18, 12].

Long queuing delay. The queuing delay is the time it takes to drain the queue of messages ahead of the tagged one. The queuing delay depends on data rate and the amount of competing traffic traversing network. In DTNs where a disconnected end-to-end path is rather common situation, the queuing time could be extremely large, e.g. minutes, hours or even days. As a result, for designing routing and forwarding mechanisms we should take into account that messages may be stored for long periods of time at intermediate nodes, and the choice of the next hop sometimes needs to be changed. The messages should be forwarded to alternative next hops if new routes to the destination are discovered [22,12].

Resource limitation. Nodes in DTNs often have very limited energy sources either because they are inherently mobile or because the power grid is non-existent in their area of location. End systems consume energy by sending, receiving, storing messages and by performing routes discovery and computation. Hence, the routing strategies sending fewer bytes and performing less computation operations are often more energy efficient [16]. In some application scenarious (e.g. wireles sensor networks) nodes are also characterized by very limited memory and processing capability [11].

Limited longevity. In some DTNs, end nodes may be deployed in hostile environments. This is especially true for sensor networks, military applications of DTNs and networks of devices used by emergency personnel [16]. In such cases, network nodes may be broken down and not be expected to last long. Recalling that the end-to-end path between two communicating entities may not exist for a long period of time, there could be the case when the delay of message delivery may exceed the lifetime of a transmitter node. As a result, the end system should not be made responsible for reliable delivery of data using classic transport layer protocols such as TCP. This feature needs to be delegated to the network [21].

Security. DTNs are vulnerable to many malicious actions and bring a number of new security challenges. The use of intermediate nodes as relays offers extraordinary opportunities for security attacks, including compromising information integrity, authenticity, user privacy and system performance. The use of specific routing mechanisms including flooding-based ones may even increase the risks associated with inserting false information into the network. Extra traffic injected by malicious nodes creates another serious threat due to resource scarcity of DTNs in some application scenarios. Unauthorized access and utilization of DTN resources for specific malicious actions are other serious concerns. It is important to note that the research on DTN security is more challenging compared to conventional mobile ad hoc networks due to its unique security characteristics. These characteristics include exceptionally long delivery delays, sporadic connectivity, opportunistic routing, and make most existing security protocols designed for conventional ad hoc networks unsuitable for DTNs [19, 22].

4. Information delivery in DTNs

The current Internet architecture and protocols are extremely successful in providing different communication services in wired and wireless networks, using TCP/IP family of protocols. However, as we discussed these set of protocols may significantly degrade performance or even disrupt operation of the network in challenged and more dynamic environments. Within the set of networking mechanisms the routing protocol is one of the main objects affecting performance of information delivery. Indeed, it is up to the routing protocol to timely discover routes in the network and maintain the uniform view of the network. During the last ten years there have been enormous research efforts trying to adapt and improve various routing protocols originally proposed for wired and wireless networks to the case of DTNs. AODV [8], DSR [10], and OLSR [33] are just few examples of routing protocols offering relatively good performance in MANETs. However, these protocols may entirely stop network activities due to intermittent connection property. As shown in the Fig. 2 the major restriction of these protocols stems from the fact that they can work and find a route only if there is an end-to-end path between end systems. Otherwise, packets will be dropped by intermediary nodes if no link to the next hop exists at the moment.

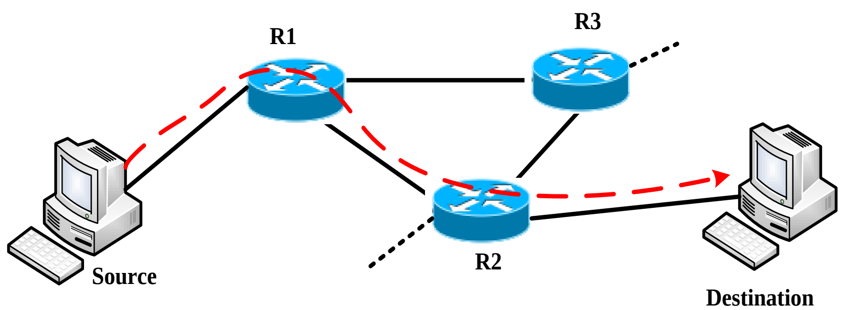

Figure 2. Illustration of routing using traditional routing protocols.

In DTNs the paths between end systems may frequently disconnect due to resource limitations, node mobility, sporadic channel availability and other DTN-specific properties. Recalling that there is absolutely no guarantee of timely delivery of data between end systems, it is unrealistic to expect that routing protocols can keep track of the topology changes in a timely manner.

One of the most important issues in a routing protocol algorithm is "routing objective". In traditional routing schemes, minimizing some metrics such as propagation delay or the number of hops in path selection process may be considered as routing objective. However, in DTNs, most of such objectives are not apparent. One of the possible metrics of interest can be message delivery to end nodes with maximum probability, while minimizing the end-to-end delay. In this section we classify routing protocols proposed for DTNs and discuss some of them.

4.1 Routing-based approaches

In DTNs the route from a source to a destination is affected by opportunity of communication between intermediate nodes. These opportunistic contacts may have time-varying and temporal properties such as capacity, rate, latency and availability. As a result, the forwarding decision, should not only take into account the number of hops between the source and the destination but also other metrics too. Links availability also is one of these metrics. The routing process becomes more complicated if link availability is nondeterministic. Utilizing knowledge about the current state and using the ability to predict the future state of the network topology may significantly improve the choice of the optimal route eventually leading to more effective forwarding of data.

Network topology in DTNs could be classified as deterministic and stochastic. In deterministic case the network topology and/or its characteristics are assumed to be known. Contrarily, for stochastic topologies no exact knowledge of topology is assumed. There are specific protocols developed for each category.

4.1.1 Deterministic routing

The main idea in computing the optimal route from a source to a destination in deterministic routing protocols is based on completely known or predictable information about nodes future mobility patterns and links availability between them. Deterministic routing protocols could be divided into the following four approaches. Most of those are special modifications of well-known algorithms.

Oracles based. Several oracle-based deterministic routing algorithms taking the advantage of predictable information about network topology and traffic characteristics have been suggested by Jain *et al. (2004).* Based on the amount of information they need to compute routes, the oracle-based algorithms are classified into complete knowledge and partial knowledge. Complete knowledge protocols utilize all information regarding traffic demands, schedules of contacts, and queuing in the forwarding process. However, in practical applications this knowledge is partially missing and routing needs to utilize available information. The authors in [30] purposed their routing framework by modifying the Dijkstra's shortest-path algorithm assuming four knowledge oracles: (i) contact summary oracle provides the knowledge about aggregated statistics of contacts, (ii) contact oracle maintains information regarding the links between two nodes at any given time, (iii) queuing oracle presenting the queuing information in each node instantaneously, and (iv) traffic demand oracle provides the knowledge about the current and future traffic characteristics. Oracle-based algorithms are mostly suitable for networks with controlled topology or with existing full or partial information about that[30,4].

Link state based. Gnawali *et al. (2005)* presented a modification of link state routing (LSR) protocol for use in deep-space networks, entitled “positional link-trajectory state” (PLS) protocol. PLS is a position-based routing mechanism that predicts the satellite or other spacecrafts moving paths to make routing decisions. In the suggested routing protocol, flooding is performed at first and then the predicted trajectory of nodes, links availability and their characteristics such as latency, error and rate through the network and link states are updated. Finally, each node independently recomputes its own routing table using a modified Dijkstra algorithm [24].

Space–time based. Merugu *et al. (2004)* suggested a routing framework, which unlike conventional routing tables using only connectivity information, provides a space-time routing table relying on information about destination and arrival time of messages. These two metrics are used to choose the next hop in a route. The underlying reason behind this approach is that in wireless networks with mobile nodes, the network topology changes with time and choosing the best route depends not only on destination but also on the topology evolution. The forwarding table in each intermediate node is a two dimensional matrix composed of destination address and instances of time when this route has been obtained. The forwarding decision is a function of both destination and time [23].

Tree based. Handorean *et al. (2005)* presented a tree-based routing algorithm based on the knowledge about motion and availability patterns of mobile nodes. Depending on how the routing information is obtained they classified the path selection mechanisms into three cases : (i) the source node initially has complete information about speed and direction of motion of all other nodes and has the ability to estimate route trees for data delivery to destination nodes, (ii) the source originally has no information about other nodes motions and each node exchanges its own information with its neighbors and learns the path to a destination whenever they meet. The second method is useful in applications where nodes have highly mobile patterns and obtaining the global knowledge is difficult (e.g. emergency networks). (iii) the future trajectory of nodes is predicted relaying on the past recorded knowledge[28]. The tree-based routing protocol requires maintenance algorithms to somehow keep the tree alive.

4.1.2 Stochastic routing: passive routing

When there is no information about nodes mobility patterns obtained via deterministic predictions or historic information stochastic routing mechanisms need to be used. Depending on whether nodes dynamically adapt their trajectories or mobility patterns to improve the routing process, stochastic routing protocols are classified into passive and active protocols.

Protocols classified as passive routing assume that the moving path of nodes does not change in order to dynamically adapt to the routing and forwarding process of messages. The basic idea of these mechanisms is to combine routing with forwarding by flooding multiple copies of a message to the network by a source and waiting for successful reception. Obviously, the more the copies of the message are available, the more the probability of the message delivery. As one can see this scheme may provide low delay at the expense of worse resource utilization. This approach is useful in those networks, where forwarding and storage resources of nodes are not scarce and there is nothing or very little knowledge regarding topology and nodes mobility [9]. In following we discuss several routing protocols using passive stochastic techniques.

Epidemic routing. Epidemic routing algorithm was the method which firstly introduced by Demers *et al.* in [4] to synchronize databases which use replication mechanism. This algorithm was modified by Vahdat *et al.(2000)* and proposed as a flooding-based forwarding algorithm for DTNs. In the epidemic routing scheme, the node receiving a message, forwards a copy of it to all nodes it encounters. Thus, the message is spread throughout the network by mobile nodes and eventually all nodes will have the same data. Although no delivery guarantees are provided, this algorithm can be seen as the best-effort approach to reach the destination. Each message and its unique identifier are saved in the node's buffer. The list of them is called the summary vector. Whenever two adjacent nodes get opportunity to communicate with each other, they exchange and compare their summary vectors to identify which messages they do not have and subsequently request them. To avoid multiple connections between the same nodes, the history of recent contacts is maintained in the nodes caches [6]. The scheme of message distributions is shown in the Fig. 3.

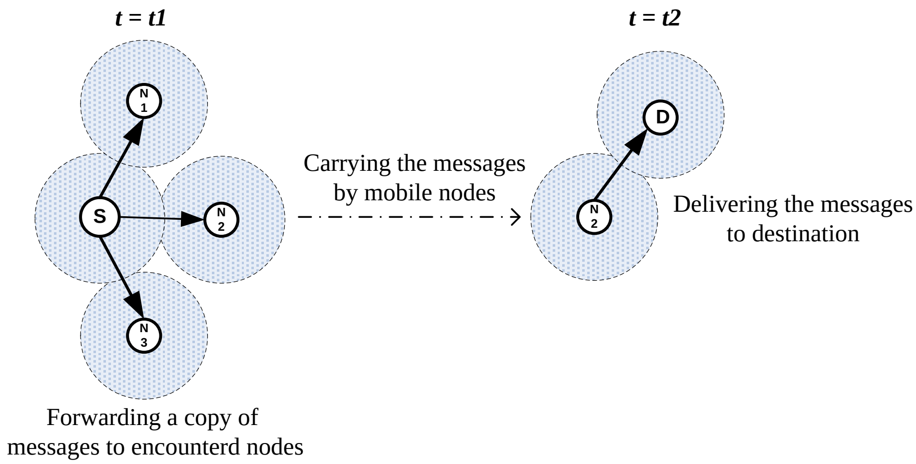

Figure 3. Illustration of epidemic routing.

Assuming sufficient resources such as node buffers and communication bandwidth between nodes, the epidemic routing protocol finds the optimal path for message delivery to destinations with the smallest delay. The reason is that the epidemic routing explores all available communication paths to deliver messages [30] and provides strong redundancy against node failures [11]. The major shortcoming is wasting of resources such as buffer, bandwidth and nodes power due to forwarding of multiple copies of the same message. It causes contentions when resources are limited, leading to dropping of messages [37]. The epidemic routing has been suggested to use in those conditions when there are no better algorithms to deliver messages. It is especially useful when there is lack of information regarding network topology and nodes mobility patterns [28].

Spray-and-wait. Wasteful resource consumption in the epidemic routing, could be significantly reduced if the level of distribution is somehow controlled. Spyropoulos *et al. (2005)* proposed the spray and wait mechanism to control the level of spreading of messages throughout the network. Similar to the epidemic routing, the spray and wait protocol assumes no knowledge of network topology and nodes mobility patterns and simply forwards multiple copies of received massages using flooding technique. The difference between this protocol and the epidemic routing scheme is that it only spreads L copies of the message. The authors in [31] proved that the minimum level of L to get the expected delay for message delivery depends on the number of nodes in the network and is independent of the network size and the range of transmission.

The spray and wait method consists of two phases, spray phase and wait phase. In the spray phase the source node after forwarding *L* copies of message to the first *L* encountered nodes, goes to the wait phase, waiting for delivery confirmation. In the wait phase all nodes that received a copy of the message wait to meet the destination node directly to deliver data to it. Once data is delivered confirmation is sent back using the same principle.

To improve the performance of the algorithm Spyropoulos *et al.(2005)* purposed the binary spray and wait scheme. This method provides the best results if all the nodes' mobility patterns in the network are independent and identically distributed (iid) with the same probability distribution. According to the binary spray and wait, the source node creates *L* copies of the original message and then, whenever the next node is encountered, hands over half of them to it and keeping the remained copies. This process is continued with other relay nodes until only one copy of the message is left. When this happens the source node waits to meet the destination directly to carry out the direct transmission [31].

In general, various methods limiting the number of distributed messages reduce recourse consumption and contention in intermediate nodes and often result better performance compared to the classic epidemic routing scheme, especially in highly loaded network conditions.

PROPHET. The probabilistic routing protocol using history of encounters and transitivity (PROPHET) is a probabilistic routing protocol developed by Lindgren *et al.* (2003). The basic assumption in the PROPHET is that mobility of nodes is not purely random, but it has a number of deterministic properties e.g. repeating behavior. In the PROPHET scheme, it is assumed that the mobile nodes tend to pass through some locations more than others, implying that passing through previously visited locations is highly probable. As a result, the nodes that met each other in the past are more likely to meet in the future [5]. The first step in this method is the estimation of a probabilistic metric called delivery predictability, $P(a,b)$Є[0,1]. This metric estimates the probability of the node *A* to be able to deliver a message to the destination node *B*. Similar to the epidemic routing, whenever a node comes in to contact with other nodes in the network, they exchange summary vectors. The difference is that in the PROPHET method the summary vectors also contain the delivery predictability values for destinations known by each node. Each node further requests messages it does not have and updates its internal delivery predictability vector to identify which node has greater delivery predictability to a given destination [5]. The operation of the PROPHET protocol could be described in two phases: calculation of delivery predictabilities and forwarding strategies.

In calculation of delivery predictabilities phase, nodes update their delivery predictability metrics whenever meet each other. Visiting more nodes results in higher delivery predictability. If two nodes do not meet each other for a long time, they exchange messages with low probabilities, so they tag their delivery predictability values as aged. Delivery predictability has transitive property meaning that if node *A* often meets node *B* and node *B* often meets node *C*, then node *A* most likely comes into contact with node *C*.

Unlike conventional routing protocols that base their forwarding decisions and selection of a path to a given destination on some simple metrics such as the shortest path or the lowest cost, forwarding strategy in the PROPHET is more complicated. Whenever a node receives a message and has no path to the destination it buffers the message and forwards it whenever another node is met. At this step, the forwarding decision could be affected by various issues. For example, forwarding more copies of the received message results in higher delivery predictability and lower delivery delays. On the other hand, more resources are spent. When a node meets a neighbor with low delivery predictability, there is no guarantee that it would meet another node with a higher delivery predictability during the message life time. As a result, a reasonable threshold must be defined for the forwarding decision [5]. Finally, it is important to mention that according to [2, 29], “PROPHET is the only DTN routing protocol that has been formally documented using RFC draft [Prophet09]”.

MobySpace. Leguay *et al. (2005)* suggested a mobility pattern space routing method called MobySpace. The major principle behind their proposal is that the two nodes with similar trajectories will meet each other with high probability. According to this method, each node forwards the received messages to the encountered nodes provided that they have similar mobility patterns with the destination node. The title of this protocol comes from a virtual Euclidean space used for taking decisions on the message forwarding process. In this virtual space each nodes is specified using its mobility pattern, called MobyPoint and routing is done towards nodes having similar MobyPoint with the destination node [14]. Each axis in the MobySpace defines the possible contact and the distance from each axis presents the communication probability between nodes. In the MobySpace the closer nodes have higher probability to communicate with each other, so in the routing process the messages are forwarded toward the nodes that are as close to the destination node as possible [13,15]. Fig. 4 illustrates forwarding paths in the MobySpace protocol.

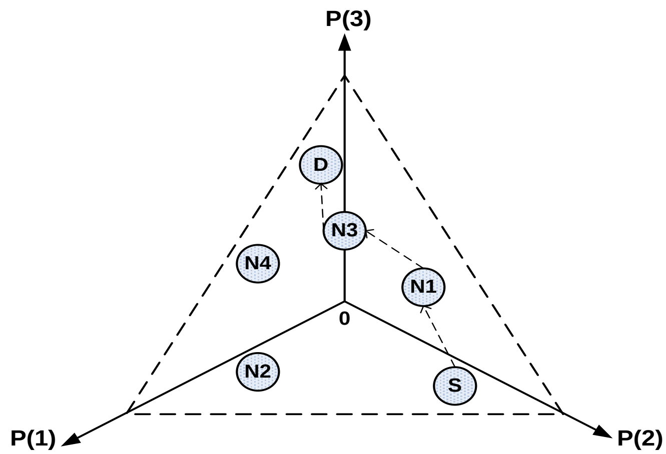

Figure 4. Forwarding path in the MobySpace.

The MobySpace protocol demonstrates better results whenever nodes' mobility patterns are fixed. However, two nodes with similar mobility patterns may never communicate if they are separated in time. In the other words, the nodes with similar trajectories could meet each other provided that they are in the same time dimension [13].

4.1.3 Stochastic routing: active routing

In this category of routing protocols moving path of some nodes are controlled in order to increase the message delivery probability. As demonstrated in the Fig. 5, in these schemes mobile nodes act as natural "message carriers" and after picking up and storing the messages from the source node move toward the destination node to deliver them. Very often the active routing methods are more complicated and costly in terms or resources that are not related to communications compared to the passive routing techniques. However, they may drastically improve the overall performance of system in terms of delay and loss metrics [37, 7]. Active routing techniques could be implemented in those DTNs where no direct communication opportunities between end systems are expected by default, e.g. emergency and military networks. Buses, unmanned aerial vehicles (UAVs) or other types of mobile nodes can be used as ferry nodes in different DTN environments [13]. In this section we discuss several routing protocols using active stochastic techniques.

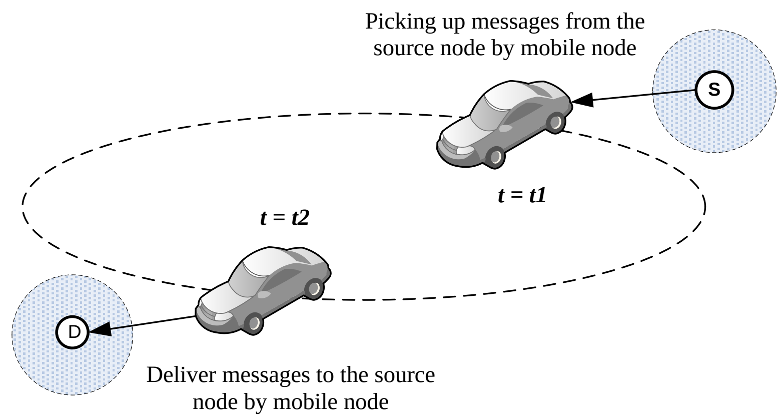

Figure 5. Message ferrying scheme in active routing protocols.

Meet-and-visit (MV). Burns *et al. (2005)* suggested the so-called meet and visit strategy for forwarding messages in structures with mobile source and fixed destination nodes. This scheme actively explores information about meeting of peer nodes and their visiting locations. The knowledge regarding meetings and visiting places is stored at each node and used to estimate message delivery probabilities. Three important assumptions are introduced in the MV protocol: (i) nodes have unlimited buffer space (ii) there is infinite link capacity (iii) and destination nodes are fixed. The mobile ground-based or airborne autonomous agents are used to provide more communication opportunities in the network. The agents trajectories are adapted according to network routing and performance requirements. The so-called autonomous controller also is responsible to control the nodes movement path according to traffic demands [7].

Message ferrying (MF). Zhao *et al. (2004)* described the so-called message ferrying method which uses mobile nodes with stable mobility patterns as collectors and carriers of messages. The ferry nodes could provide connectivity among nodes in a network, where there are no possibilities for direct communication between end systems. Because of fixed moving path of ferry nodes, each node can save information about the ferries' mobility patterns and may adapt its future trajectory to come into contact with ferry nodes to have sending or receiving messages. Depending on the entity initiating transactions, two forwarding schemes can be used for message delivery: node-initiated message ferrying (NIMF) and ferry-initiated message ferrying (FIMF). According to the first approach the ferry nodes choose their path using a predefined mobility pattern known by other nodes. Whenever the nodes want to send messages via the ferries, they need to adjust their trajectories to move towards the ferry nodes. The nodes can be informed about ferries' paths using broadcasting messages originated by ferry nodes or using pre-defined schedules. In the FIMF, nodes broadcast call-for-service requests whenever they need to send or receive messages. The nearest ferry node is responsible for responding them and moving towards the nodes to pick up the messages [36].

4.2 Network Coding based techniques

The idea of network coding-based information delivery has been introduced in the seminal paper by Ahlswede *et al* [27]. Network coding was originally developed to increase capacity of wired networks operating in multicast mode.

4.2.2 Network Coding Advantages

Network coding concept allows intermediate relay nodes act as a special coder mixing incoming packets instead of simply replicating them to one or a set of output ports. This scheme illustrated in Fig. 6 using the well known butterfly network example. In this example, node *N3* after *XORing* the *X* and *Y* messages, forwards it to end nodes through node *N4*. Then the nodes *N5* and *N6* respectively can decode *Y* and *X* by *XORing* the received messages. Using *XOR* coding in this network allows channels to be used just once. Without network coding, the channels between the nodes *N3* and *N4* to end nodes must be used twice [27].

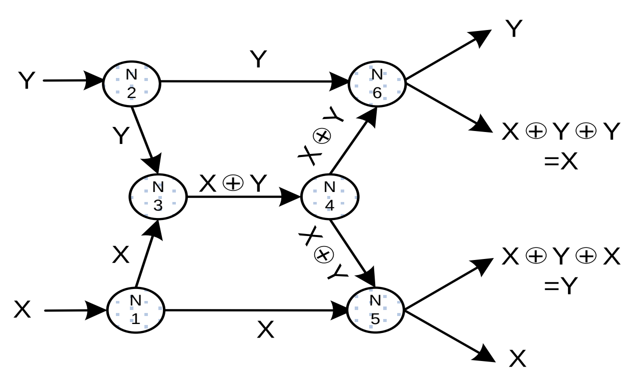

Figure 6. Examples of network coding in multicasting link.

In networks without intermediate coding, destination nodes need to receive specific number of successively sent packets by the source node to determine information completely. As it is demonstrated in Fig. 7, usage of network coding provides the receiver the ability to decode and exploit all sent information by receiving a reasonable number of independent encoded packets. So the lossy network could be more reliable by using network coding mechanism [17].

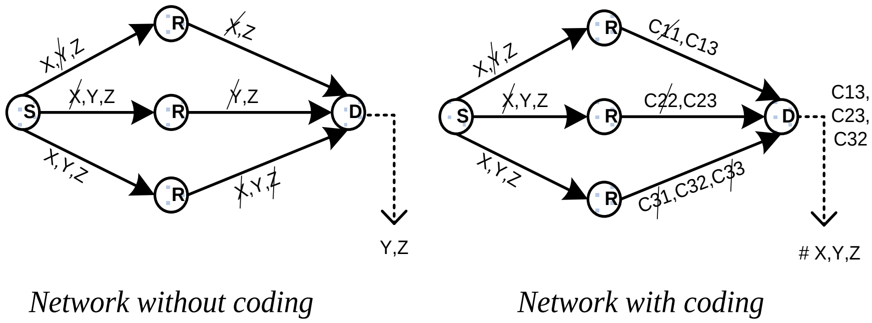

Figure 7. Improving reliability using network coding.

Benefits of network coding eventually led to adopt this idea for broadcast-based wireless networks, where nodes are often subject to resource limitations in terms of power, buffer and link capacity. In various wireless networks such as sensor, mesh, vehicular networks and DTNs links between end systems are inherently intermittent due to dynamic network topology. To enable efficient communication, intermediate mobile or stationary nodes are responsible for acting as relays using the store-and-forward mechanism. If the buffer of a node is filled up and new data arrives before delivery of the stored messages, the node may drop them or delete the old messages to store new ones. As shown in Fig. 8, usage of network coding makes possible to mix and code newly arrived and old data in buffer and generate new encoding vectors as a function of all received data, without deleting any data in the buffer or dropping new ones [40].

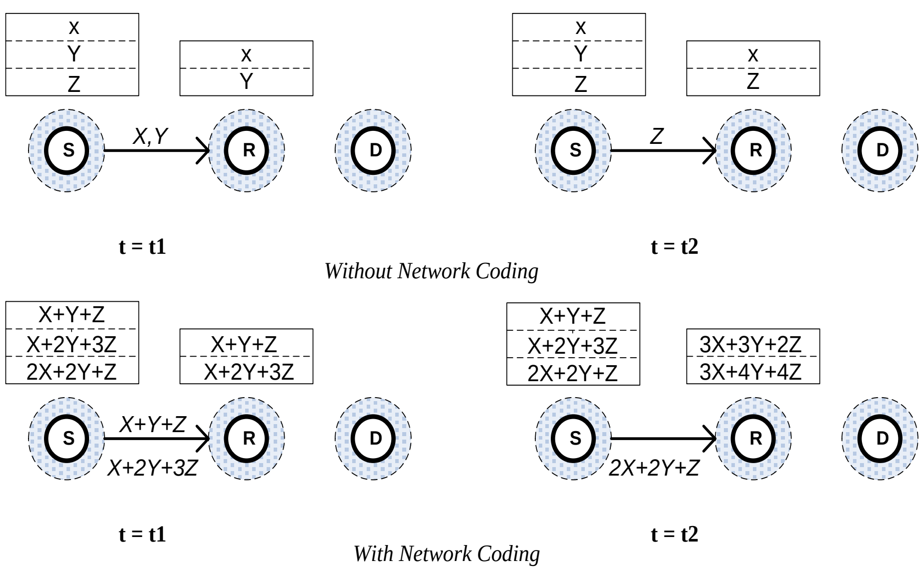

Figure 8. Improving buffer performance using network coding.

Implementing network coding in network nodes imposes additional processing overhead due to encoding at intermediate nodes and decoding at the destination. More complex coding offers better performance at the expense of higher processing overhead. As a result, in those environments where processing power is a scarce resource simple network coding algorithms like *XOR* or linear methods could be used.

XOR coding. In this method, the intermediate relay nodes broadcast *XORed* version of messages to all nodes after receiving them from the source nodes. Corresponding destination nodes should be able to decode the sent messages by *XORing* the received messages with themselves. It is important to note that only those nodes knowing one of two elements of the encrypted messages can recover the sent messages [15, 37]. The latter property improves security of wireless transmission.

Linear coding. As the title suggests, mixing of the received messages at intermediate nodes is done using linear combinations. The coefficients of this combination are taken from a finite field that needs to be the same for all nodes. If the received packets and their combinations are denoted by x_i and g_i respectively, the linear combination of the packets is given by $\sum_i g_i \times x_i$. The destination nodes can decrypt the encrypted data by receiving *n* combinations of *N* sent messages provided that the rank of combinations equals to *n*. Receiving more combinations of messages results in higher probability of correct decoding, only if the coefficients are set randomly at intermediate nodes [32, 26]. Random linear network coding could be used to achieve that. In such mechanism each node combines input packets using random coefficients in a random linear manner [6]. Gaussian elimination algorithm is used by destination nodes to solve the matrix with *n* equations to retrieve *N* unknown parameters that represent the sent messages [32].

4.2.3 Network coding based routing method

Network coding-based routing is an adaptation of the traditional store-and-forward mechanism. According to it, relay nodes combine and encode received packets before forwarding them. This approach exploits the basic principle of network-coding consisting in limiting the number of message forwarding in resource restrictive conditions. In traditional replication-based routing methods, successful transmission is achieved by delivering each of data packets separately; while in the network coding-based scheme destination nodes can recover sent packets by receiving only a reasonable number of coded packets. It means that coding-based schemes result in more reliable communications in poor and resource limited connectivity conditions. However imposing additional computing overhead at nodes due to coding and decoding processes causes to performance degradation in networks with stable and well connected links [20]. Due to having outstanding benefits, network coding is currently one of the most promising research areas within the scope of information delivery in DTN-like networks. In this section we discuss two different proposals based on network-coding.

Network coding based epidemic routing (NCER). Lin *et al. (2007)* developed a protocol based on combination of the network coding and epidemic routing. According to this protocol, instead of replicating just copies of messages as used in old epidemic routing, relay nodes flood random linearly coded versions of the received packets to other nodes whenever they have an opportunity to communicate. It has been shown that the suggested method provides better performance comparing traditional epidemic routing protocol in terms of average delivery delay, storage and bandwidth usage. More information about this protocol can be found in [41].

Efficient routing protocol based on network coding (E-NCP). E-NCP is another network coding-based routing protocol presented by Lin *et al. (2008)*. Recall that in NCER nodes forward encoded messages until the successful delivery is signaled by acknowledgement packet or the message lifetime expires. E-NCP protocol is intended to improve NCER performance by optimizing the number of forwarded messages. Based on information theory, destination nodes should receive at least N randomly encoded packets in order to recover N original transmitted data packets with probability 1. In E-NCP in order to deliver data with high probability, the source node starts by sending more than N encoded packets spreading them to L encountered relay nodes using the binary spraying scheme. Messages delivery delay and the number of relay nodes can be controlled by adjusting the L. In [42] the authors carried out detailed evaluation and optimization of the protocol.

5. Conclusions

DTNs are promising approaches to enable communications in those networks offering no continual end-to-end communication links between nodes. Routing and forwarding is one of the important key points that considerably affect the overall performance of networks. In this article we have discussed various routing protocols proposed for DTNs and categorized existing ones into two basic classes. These are deterministic and stochastic routing protocols. Classification is based on making forwarding decision in routing methods with and without the knowledge about network topology and nodes trajectories. Protocols in each class have their own advantages and shortcomings. Deterministic routing methods often are more complex compared to stochastic routing protocols. However, they provide better performance in networks where there are information regarding network topology and nodes mobility patterns. More knowledge results in effective message delivery methods and efficient resource consumption. Simple flooding based routing protocols is a feasible approach in those networks, where there is a little or no information existent about network topology and network is without resource restrictions. Improvement of capacity, reliability and performance in wired and wireless networks are results of network coding paradigm. Usage of network coding-based routing protocols to improve efficiency of DTNs is one of the novel fields presented in the last section of this paper. This survey could be helpful to those interested in information delivery in DTN systems and trying to get the overall knowledge regarding the state-of-the-art routing schemes in delay tolerant networks.

References

[1] “Delay Tolerant Networking,” http://www.nasa.gov/mission_pages/station/research/.

[2] A. Lindgren, A. Doria, “Probabilistic routing protocol for intermittently connected networks,” IETF draft, 2002.

[3] United Villages project, http://www.unitedvillages.com/

[4] A. Demers, D. Greene, C. Houser, W. Irish, J. Larson,S. Shenker, H. Sturgis, D. Swinehart, and D. Terry, “Epidemic algorithms for replicated database maintenance,” ACM SIGOPS Operating Systems Review, V.22, N.1, Jan. 1988.

[5] A. Lindgren, A. Doria, O. Schelen, “Probabilistic routing in intermittently connected networks,” ACM SIGMobile Computing and Communication Review, V.7, N.1, July 2003.

[6] A. Vahdatand, D. Becker, “Epidemic Routing for Partially-Connected AdHoc Networks,” Duke Technical Report, CS-2000-06, available at issg.cs.duke.edu/epidemic/epidemic.pdf, accessed on 06.03.2011, Jul. 2000,

[7] B. Burns, O. Brock, and B.N. Levine, “MV routing and capacity building in disruption tolerant networks”, In Proc. IEEE INFOCOM, pp. 398-408, March 2005.

[8] C. Perkins, “Ad hoc on demand distance vector (AODV) routing,” IETF Draft, 1997.

[9] D. Choi “Challenges and Applications of Delay Tolerant Networks,” ECE, University of Waterloo, available on bbcrlab-pc9.bbcrlabpcnet.uwaterloo.ca, accessed on 11.02.2011.

[10] D. B Johnson and D. A Maltz. "Dynamic source routing in ad hoc wireless networks," Mobile Computing," V. 353, Ch.5, Kluwer Academic Publishers, 1996.
[11] Evan P.C. Jones and Paul A.S. Ward, "Routing Strategies for Delay-Tolerant Networks," 2006, available at citeseerx.ist.psu.edu, accessed on 05.02.2011.
[12] Evan P. C. Jones, Lily Li, Jakub K. Schmidtke, Paul A. S. Ward, "Practical Routing in Delay-Tolerant Networks," IEEE Trans. Mob. Comput., V.6, N.8, pp. 943-959, Aug 2007.
[13] I. Cardei, C. Liu, J. Wu, "Routing in Wireless Networks with Intermittent Connectivity," Encyclopedia of Wireless and Mobile Communications, B. Furht (ed.), CRC Press, Taylor and Francis Group, 2007.
[14] Jeremie Leguay, Timur Friedman, Vania Conan, "Evaluating Mobility Pattern Space Routing for DTNs," In Proc. IEEE INFOCOM, pp. 1-10, Apr. 2006
[15] Jian Zhang, Yuanzhu Peter Chen, Ivan Marsic, "Network Coding via Opportunistic Forwarding in Wireless Mesh Networks," In Proc. WCNC, pp. 1775-1780, 2008.
[16] J. Shen, S. Moh, and I. Chung, "Routing Protocols in Delay Tolerant. Networks: A Comparative Survey", In Proc. 23rd IEICE ITC-CSCC, pp. 1577-1580, Aug 2007.
[17] J. Widmer and J.-Y. L. Boudec, "Network coding for efficient communication in extreme networks," In Proc. of the ACM SIGCOMM, Workshop on DTNs, 2005.
[18] A. Haris, "A DTN study analysis of implementations and tools," Master's Thesis, TKK/HUT, 2010, available at http://nordsecmob.tkk.fi/Thesisworks/Abdullah20Haris.pdf, accessed on 15.02.2011.
[19] Haojin Zhu, "Security in Delay Tolerant Networks," PhD thesis, ECE, University of Waterloo, 2009, available at http://uwspace.uwaterloo.ca/bitstream/10012/4348/, accessed on 18.02.2011.
[20] H. Nguyen, S. Giordano, "Routing in Opportunistic Networks," International Journal of Ambient Computing and Intelligence (IJAC), V.1, N.3, pp. 19-38, 2009.
[21] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets," In Proc. ACM SIGCOMM, Feb. 2003
[22] L.Pelusi, A.Passarella, M. Conti, "Opportunistic networking: data forwarding in disconnected mobile ad hoc networks," IEEE Com. Mag., V.44, N.11, pp. 134-141, 2006.
[23] Merugu, S., Ammar, M., Zegura, E., "Routing in Space and Time in Networks with Predictable Mobility," Tech. Report, GIT-CC-04-7, Georgia Institute of Tech., 2004, available at http://smartech.gatech.edu/handle/1853/6492, accessed on 01.03.2011.
[24] O. Gnawali, M. Polyakov, P. Bose, R. Govindan, "Data centric, position-based routing in space networks," In Proc. 26th IEEE Aerospace Conference, pp. 1322-1334, 2005.
[25] M. Pierobon, I. Akyildiz, "A Physical End-to-End Model for Molecular Communication in NanoNetworks," IEEE JSAC, V.28, N.4, pp. 602-611, May 2010.
[26] R. Yeung, S.-Y. Li, N. Cai, Z. Zhang, "Network Coding Theory," Foundations and Trends in Communications and Information Theory, V.2, N.5, NowPublishers, 2005.
[27] R. Ahlswede, N. Cai, S.-Y. Li, R. Yeung, "Network Information Flow," IEEE Transactions on Information Theory, V.46, N.4, pp. 1204-1216, July 2000.

[28] R. Handorean, C. Gill, G.-C. Roman, "Accommodating transient connectivity in ad hoc and mobile settings," LNCS, V. 3001/2004, Springer, pp. 305-322, Jan. 2004.
[29] "D2.1 DTN - The State of the Art," FP7 Project "Networking for communications challenged communities: architecture, testbeds and innovative alliances", available on www.n4c.eu/Download/n4c-wp2-012-state-of-the-art-101.pdf, accessed on 15.01.2011.
[30] S. Jain, K. Fall, and R. Patra, "Routing in delay tolerant networks," In Proc. ACM SIGCOMM, 2004.
[31] T. Spyropoulos, K. Psounis, C. Raghavendra, "Spray-and-Wait: Efficient routing scheme for intermittently connected mobile networks," in ACM SIGCOMM Workshop on Delay Tolerant Networking (WDTN), 2005.
[32] S.-Y. Li, R. Yeung, N. Cai, "Linear network coding," IEEE Transactions on Information Theory, V.49, N.2, pp. 371-381, Feb. 2003.
[33] T. Clausen, P.Jacquet, "Optimized link state routing protocol (OLSR)," RFC 3626, IETF, 2003.
[34] W. Zhao, M. Ammar, E. Zegura, "A Message Ferrying Approach for Data Delivery in Sparse Mobile Ad Hoc Networks," In Proc. Of 5th MobiHoc, pp. 113-117, 2004.
[35] W. Zhao, M. Ammar, E. Zegura., "Controlling the mobility of multiple data transport ferries in a delay-tolerant network," In Proc. IEEE INFOCOM, V.2, pp. 1407-1418, March 2005.
[36] Y.-C. Tseng, S.-Y. Ni, Y.-S. Chen, J.-P. Sheu, "The broadcast storm problem in a mobile ad hoc network," Wireless networks, V.8, N.2-3, pp. 153-167, 2002.
[37] Y. Lin, B. Li, B. Liang., "Efficient Network Coded Data Transmissions in Disruption Tolerant Networks," In Proc. IEEE INFOCOM, pp. 1508-1516, Apr. 2008.
[38] Y. Lin, B. Liang, B. Li, "Performance Modeling of Network Coding in Epidemic Routing," In Proc. MobiOpp 2007, pp. 345-351, June 2007.